Asymptotic Information Measures Discrimination of Non-Stationary Time Series Based on Wavelet Domain

Discriminacion de medidas de información asintótica de series de tiempo no estacionarias basadas en dominio wavelet

Behzad Mansouri[1,a], Rahim Chinipardaz[1,b]

[1]Department of Statistics, Faculty of Mathematics and Computer Science, University of Shahid Chamran, Ahvaz, Iran

---

Abstract

This article is concerned with the problem of discrimination between two classes of locally stationary time series based on minimum discrimination information. We view the observed signals as realizations of Gaussian locally stationary wavelet (LSW) processes. The asymptotic Kullback - Leibler discrimination information and Chernoff discrimination information are developed as discriminant criteria for LSW processes. The simulation study showed that our procedure performs as well as other procedures and in some cases better than some other classification methods. Applications to classifying real data show the usefulness of our discriminant criteria.

Key words: Chernoff information, discrimination, evolutionary wavelet spectrum, Kullback - Leibler information, locally stationary wavelet processes, seismic data.

Resumen

Este artículo se refiere al problema de discriminación entre dos clases de series de tiempo estacionarias locales basadas en información de discriminación mínima. Se consideran las señales observadas como realizaciones de procesos wavelet estacionarios locales (LSW, por sus siglas en inglés) gausianos. La información de discriminación Kullback - Leibler asintótica y la información de discriminación de Chernoff se desarrollan como criterios discriminantes para procesos LSW. El estudio de simulación mostró que el procedimiento propuesto se desempeña tan bien como otros procedimientos y en algunos casos mejor que otros métodos de clasificación. Aplicaciones

---

[a]Assistant Professor. E-mail: b.mansouri@scu.ac.ir

[b]Professor. E-mail: chinipardaz_r@scu.ac.ir

a la clasificación de datos sísmicos muestran la utilidad de los criterios discriminantes propuestos.

Palabras clave: LaTeX Datos sísmicos, discriminación, espectros wavelet evolucionarioinformación de Chernoff, información de Kullback-Leibler, procesos wavelet estacionarios locales .

1. Introduction

Many practical problems in time series analysis are reduced to classifying a stochastic process to one or a other category. Shumway & Stoffer (2011) is a good source for examples in this field. Kullback & Leibler (1951) and Chernoff (1952) information measures, hereafter KL and CH information measures, respectively, are appropriate measures for classification and discrimination of time series data. KL and CH information measures have been widely used for many years in various fields such as economics, engineering and medical sciences.Gersch, Yeager, Diamond, Spire, Gerry & Gevins (1975) and Gersch & Yonemoto (1977) used KL information measures to determine whether or not patients are sufficiency anesthetized for surgery. Shumway & Unger (1974) and Dargahi-Noubary & Laycock (1981) obtained spectral forms of KL discrimination information to distinguish between different classes of seismic data, earthquakes and explosions. In an extension of Shumway & Unger (1974), Kakizawa, Shumway & Taniguchi (1998) developed KL along with Chernoff discrimination measures to bivariate time series data. Parzen (1990) and Zhang & Taniguchi (1992) studied the robustness of KL to non-Gaussian departure. All of the afore mentioned papers are based on stationary assumption. These methods are performed admirably for stationary processes, however time series often have time-varying dynamics, and traditional spectral procedures are not performed well. Recently, to model non-stationary time series, authors have proposed several approaches such as locally stationary modeling by Dahlhaus (1997) or smooth localized complex exponential (SLEX) methodology by Ombao, Raz, von Saches & Malow (2001) and Ombao, Raz, von Saches & Guo (2002). These methods provide a better basis for the discrimination of non-stationary time series such as seismic data. Recognizing this, Shumway (2003) and Sakiyama & Taniguchi (2004) for Dahlhaus' locally stationary processes and Huang, Ombao & Stoffer (2004) for stationary time series in the SLEX model, proposed classification techniques. In the remainder of the article, we refer to these procedures as the SST and SLEX methods respectively.

Some authors like Maharaj & Alonso (2007) and Fryzlewicz (2003) have proposed methods based on wavelets to discriminate analysis of time series and employ it to discriminate seismic signals. Modeling a non-stationary random process, such as a LSW processes is discussed in detail by Nason, von Sachs & Kroisandt (2000). The LSW model provides a time-scale decomposition of the signals in which we can define and rigorously estimate the evolutionary wavelet spectrum Fryzlewicz & Ombao (2009). In this paper we view the observed signals as realizations of Gaussian LSW processes. We obtain likelihood and calculating the asymptotic KL and CH discrimination measures as discriminant criteria for Gaussian LSW processes.

Our discriminant criteria are related to evolutionary wavelet spectrum (EWS), which contains the second-moment information on the non-stationary time series. Throughout the paper we suppose that there are two different groups of Gaussian LSW processes and there is a new non-stationary time series vector that has to be allocated to one of them. For each time series to be classified, we compute the wavelet discrimination information measures and the divergence from the wavelet discrimination information measures of each group, which is then assigned to the group to which it is the least dissimilar. We recognize that our methodology is very close to the method mentioned in Fryzlewicz & Ombao (2009) (that we refer to as the LSW method), but ours differs in that the LSW method is based on the L_2 criteria while our method is based on much familiar criteria in discrimination i. e. the Kullback-Leibler and Chernoff discrimination measures.

The article is organized in six Sections as follows: In Section 2, we discuss the LSW model as a tool to analyze non-stationary time series data. In Section 3, we obtain the minimum discrimination information for two criteria, Kullback-Leibler and Chernoff discrimination measures for Gaussian LSW processes. We explain our discrimination algorithm in Section 4. To evaluate the performance of our discriminant criteria, in Section 5, a simulation study is carried out and our methods are compared with other proposed approaches. Finally, in Section 6, we apply our procedure to real data.

2. The LSW Model

Throughout the paper we assume that the non-stationary time series follows the LSW model based on the use of the discrete wavelet transform (DWT), use Nason et al. (2000). A class of LSW processes, is a sequence of doubly-indexed stochastic processes $\{X_{t,T}\}_{t=1,\ldots,T}, T \geq 1$, with the following representation in the mean-square sense in respect to a given wavelet basis $\{\psi_{j,k}(t)\}_{j,k}$ of $L_2(\mathbf{R})$, and an orthonormal random sequence of increments $\xi_{j,k}$

$$X_{t,T} = \sum_{j=-J}^{-1} \sum_{k} w^0_{j,k:T} \psi_{j,k}(t) \xi_{j,k}, \tag{1}$$

where $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), j = -1, -2, \ldots, -J(T) = -\log_2 T, k \in \mathbf{Z}$, and possess the following properties:

1. $E(\xi j, k) = 0$ for all j, k and hence $E(X_{t,T}) = 0$ for all t and T.

2. $Cov(\xi_{j,k}, \xi_{l,m}) = \delta_{j,l}\delta_{k,m}$ where $\delta_{i,j}$ is the Dirac function.

3. For each $j \leq -1$ there is a Lipschitz-continuous function $W_j(z)$ on $(0,1)$ so as

$$\sup_{k} \left| w^0_{j,k;T} - W_j\left(\frac{k2^{-j}}{T}\right) \right| = O(T^{-1}) \quad \text{for } T \to \infty,$$

where for each $j = -1, -2, \ldots, -J$ the sup is over $k = 1, \ldots, N_j = 2^{-j}$.

4. The real-valued wavelet basis $\{\psi_{j,k}(t)\}_{j,k}$ is orthonormal and the wavelets have compact support.

5. $\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty$ for all $z \in (0,1)$.

***Note* 1.** Let $S_j(z) := |W_j(z)|^2, z \in (0,1)$, then $\{S_j(z)\}_{j \leq -1}$ is the evolutionary wavelet spectrum (EWS) of the sequence $\{\mathbf{X}_{t,T}\}$ with respect to $\{\psi_{j,k}\}$. As such, Nason et al. (2000), show the following results:

1. $S_j(z) = \lim_{T \to \infty} \left| w_{j,2^j[zT];T}^0 \right|^2, \forall z \in (0,1)$.

2. $\lim_{T \to \infty} Var\{\mathbf{X}_{[zT],T}\} = \sum_{j=-\infty}^{-1} S_j(z) < \infty, \forall z \in (0,1).$ \hfill (2)

3. $\lim_{T \to \infty} Cov\{\mathbf{X}_{[zT]-\tau,T}; \mathbf{X}_{[zT]+\tau,T}\} = \sum_{j=-J}^{-1} S_j(z)\Psi_j(\tau), \tau \in Z, z \in (0,1),$ \hfill (3)

where $Cov\{\mathbf{X}_{[zT]-\tau,T}; \mathbf{X}_{[zT]+\tau,T}\}$ and $\Psi_j(\tau) = \sum_k \psi_{j,k}(0)\psi_{j,k}(\tau), j < 0$ are local autocovariance and autocorrelation wavelets, respectively.

***Note* 2.** Note that most of the time series, including the stationary series can be modeled as (1).

3. Kullback-Leibler and Chernoff Information Measures

3.1. Discrimination Based on Using Decimated Wavelet Basis

Let $d_{j,k:T} = \sum_{t=1}^{T} \mathbf{X}_{t,T}\psi_{j,k}(t)$ be the empirical wavelet coefficients, where $\mathbf{X}_{t,T}$ is a LSW process given by (1). Then:

$$d_{j,k:T} = \sum_{t=1}^{T} \mathbf{X}_{t,T}\psi_{j,k}(t) = \sum_{t=1}^{T} \left\{ \sum_l \sum_m w_{l,m;T}^0 \psi_{lm}(t)\xi_{l,m} \right\} \psi_{j,k}(t)$$

$$= \sum_l \sum_m w_{l,m;T}^0 \xi_{lm} \left\{ \sum_{t=1}^{T} \psi_{l,m}(t)\psi_{j,k}(t) \right\} = w_{j,k;T}^0 \xi_{j,k}.$$

Hence, since the $\xi_{l,m}$ are uncorrelated:

$$Cov(d_{j,k;T}, d_{j',k';T}) = ||w_{j,k;T}^0||^2 \delta_{j,j'} \delta_{k,k'}.$$

If $\xi_{j,k}$ in (1) is a Gaussian process, the empirical wavelet coefficients are independent Gaussian variables, i.e:

$$d_{j,k:T} \sim N(0, |w_{j,k:T}^0|^2) \hfill (4)$$

then, using formula (4), the log likelihood ratio based on empirical wavelet coefficients becomes

$$LLR = -\frac{1}{2}\ln\left(\frac{\prod_j \prod_k |w^0_{1,j,k;T}|^2}{\prod_j \prod_k |w^0_{2,j,k;T}|^2}\right)$$
$$+\frac{1}{2}\sum_{j=-1}^{-J}\sum_{k=1}^{N_j}\left(\frac{1}{|w^0_{2,j,k;T}|^2}-\frac{1}{|w^0_{1,j,k;T}|^2}\right)d^2_{j,k;T} \quad (5)$$

Consider the situation where a LSW process $X_{t,T}$, under H_i belong to population Π_i with representation $X_{t,T} = \sum_{j=-J}^{-1}\sum_k w^0_{i,j,k;T}\psi_{j,k}(t)\xi_{j,k}$ for $i = 1, 2$. One classical measure for disparity between two multivariate distribution is the Kullback-Leibler (KL) discriminant information (Kullback & Leibler 1951, Kullback 1978), defined as

$$I(1:2) = E_{H_1}\left\{\ln\frac{p_{\Pi_1}}{p_{\Pi_2}}\right\} \quad (6)$$

where p_{Π_i} is the probability density of $X_{t,T}$, under $H_i, (i = 1, 2)$.

Another criterion for measuring the discrepancy between two densities, proposed in Parzen (1990), is Chernoff information based, which is on Chernoff distance (see Chernoff 1952) and given by

$$Q_h(1:2) = -\frac{1}{2}\ln E_{H_1}\left\{\left(\frac{p_{\Pi_2}}{p_{\Pi_1}}\right)^h\right\} \quad (7)$$

where $h \in (0, 1)$.

Using (4) and (5), KL criterion and Chernoff criterion can be obtained as

$$I(1:2) = \frac{1}{T}E_{H_1}\{LLR\} = \frac{1}{2}\sum_{j=-1}^{-J}\sum_{k=1}^{N_j}\left[\frac{|w^0_{1,j,k;T}|^2}{|w^0_{2,j,k;T}|^2}+\frac{1}{2}\ln\frac{|w^0_{2,j,k;T}|^2}{w^0_{1,j,k;T}|^2}-1\right] \quad (8)$$

and

$$Q(1:2) = -\frac{1}{T}\ln\{E_{H_1}\{(LLR)^h\}$$
$$= -\frac{1}{T}\left\{\ln\left[\frac{\prod_j\prod_k|w^0_{1,j,k;T}|^2}{\prod_j\prod_k|w^0_{2,j,k;T}|^2}\right]^{\frac{h}{2}}+\ln\left[\prod_{j=-1}^{-J}\prod_{k=1}^{N_j}\left(1-h\left(1-\frac{|w^0_{1,j,k;T}|^2}{|w^0_{2,j,k;T}|^2}\right)\right)^{\frac{1}{2}}\right]\right\} \quad (9)$$
$$= \frac{1}{2T}\left\{\sum_{j=-1}^{-J}\sum_{k=1}^{N_j}\left[\ln\left[\frac{(1-h)|w^0_{2,j,k;T}|^2+h|w^0_{1,j,k;T}|^2}{|w^0_{2,j,k;T}|^2}\right]-h\ln\left[\frac{|w^0_{1,j,k;T}|^2}{w^0_{2,j,k;T}}\right]\right]\right\}$$

3.2. Discrimination Based on Using Non-Decimated Wavelet Basis

In an LSW model based on using non-decimated wavelet basis $\psi_{jk}(t) = 2^{j/2}\psi(2^j(t-k))$:

$$
\begin{aligned}
d'_{j,k;T} &= \sum_{t=1}^{T} X_{t,T}\psi_{j,k}(t) \\
&= \sum_{t=1}^{T}\left\{\sum_l\sum_m w^0_{l,m;T}\psi_{l,m}\right\}\psi_{j,k}(t) = \sum_l\sum_m w^0_{l,m;T}\xi_{l,m}\left\{\sum_{t=1}^{T}\psi_{l,m}(t)\psi_{j,k}(t)\right\} \\
&= \sum_l\sum_n w^0_{l,n+k;T}\xi_{l,n+k}\left\{\sum_{t=1}^{T}\psi_{l,n+k-t}\psi_{j,k-t}\right\}.
\end{aligned}
$$

Let $\xi_{j,k}$ be a Gaussian process, then

$$
\begin{aligned}
Cov(d'_{j,k;T}, d'_{j',k';T}) &= \sum_l\sum_m w^0_{l,n+k;T}\left\{\sum_{t=1}^{T}\psi_{l,n-t}\psi_{j,-t}\right\} \\
&\times \sum_v\sum_m w^0_{v,m+k;T}\left\{\sum_{s=1}^{T}\psi_{v,m-s}\psi_{j,-s}\right\}Cov(\xi_{l,n+k}, \xi_{l,m+k'}).
\end{aligned}
$$

Let $l = v$ and $n + k = m + k'$. Using Remark 1 we have

$$
\begin{aligned}
Cov(d'_{j,k;T}, d'_{j',k';T}) &= \sum_l\sum_m |w^0_{l,n+k;T}|^2\sum_t\sum_s\psi_{l,n-t}\psi_{j,-t}\psi_{l,n+(k-k')-s}\psi_{j',-s} \\
&= \sum_l S_l(k/T)\sum_n\sum_t\sum_s\psi_{l,n-t}\psi_{j,-t}\psi_{l,n+(k-k')-s}\psi_{j',-s} + O(T^{-1})
\end{aligned}
$$

let $v = s - t$, then

$$
Cov(d'_{j,k:T}, d'_{j',k':T}) = \sum_l S_l(k/T)\sum_v \Psi_l(v+(k'-k))\sum_t\psi_{j,-t}\psi_{j',-v-t} + O(T^{-1}) \tag{10}
$$

In (10) if $j' = j$ and $k = k'$ then

$$
\begin{aligned}
Var(d'_{j,k;T}) &= \sum_l S_l(k/T)\sum_v \Psi_l(v)\Psi_j(v) + O(T^{-1}) \\
&= \sum_t A_{j,l}S_{l'}(z) + O(T^{-1}) \tag{11}
\end{aligned}
$$

where

$$
A_{j,l} = \sum_\tau \Psi_j(\tau)\Psi_l(\tau). \tag{12}
$$

The covariance between two empirical wavelet coefficients $d'_{j,k;T}$ and $d'_{j',k';T}$ decays with an increasing distance between the two relative positions of the location k on

scale j and k' on scale j'. For example, when $j' = j$ and $k' \neq k$, we have

$$Cov(d'_{j,k;T}, d'_{j',k';T}) = \sum_l S_l(k/T) \sum_v \Psi_l(v + (k' - k))\Psi_j(v) + O(T^{-1}).$$

Note that $\Psi_i(v+(k'-k))$ is zero as $v+|k'-k|$ exceeds the supports of the wavelets. Therefore, the correlation between $d'_{j,k;T}$ and $d'_{j,k';T}$ vanishes as $|k' - k| \to \infty$.

As noted in Remark 1, $|w^0_{j,2^j[zT];T}|^2 = S_j(z) + O(T^{-1})$. By replacing $|w^0_{i,j,k;T}|$ by $S^i_j(k/T)$ ($for\,i = 1, 2$), in the KL criterion in (8) and CH criterion in (9), we have:

$$\hat{I}(1:2) = \frac{1}{2} \sum_{j=-1}^{-J} \sum_{k=1}^{T} \left[\frac{S^1_j(k/T)}{S^2_j(k/T)} + \ln \frac{S^2_j(k/T)}{S^1_j(k/T)} - 1 \right] \tag{13}$$

and

$$\hat{Q}(1:2) = \frac{1}{2T} \left\{ \sum_{j=-1}^{-J} \sum_{k=1}^{T} \left[\ln \left[\frac{(1-h)S^2_j(k/T) + hS^1_j(k/T)}{S^2_j(k/T)} \right] - h \ln \left[\frac{S^1_j(k/T)}{S^2_j(k/T)} \right] \right] \right\} \tag{14}$$

In practice the true wavelet spectra $S^i_j(z)$ are unknown, and should be estimated. Let $I^j_{k,T} = |d'_{j,k;T}|^2$ be the wavelet periodogram of $\{X_{t,T}\}$. From (11), $I^j_{k,T}$ is a biased estimator of the evolutionary wavelet spectrum (EWS). (Nason et al. 2000) have shown that the J-dimensional matrix $A_j = (A_{j,l})_{j,l=-1,\ldots,-J}$ is positive-definite and has a bounded inverse. Formula (11) suggests that the natural unbiased estimator of $S_j(k/T)$ is an empirical wavelet spectrum given by

$$L^j_{k,T} = \sum (A^{-1})_{i,j} I^i_{k,T} \tag{15}$$

4. Discrimination Based on KL and CH Measures

Suppose that the observed vector of time series $\mathbf{W} = \{W_{t,T}\}_{t=1,\ldots,T}.T \geq 1$ has to be assigned to one of two available groups $\{\mathbf{X}_1, \ldots, \mathbf{X}_{n_1}\}$ or $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2}\}$ with evolutionary wavelet spectra $S^1_j(z)$ and $S^2_j(z)$, respectively. The classification algorithm proceeds as follows.

1. Choice of the Discriminating Set Θ. The important wavelet coefficients for classification are identified using the method proposed by Fryzlewicz & Ombao (2009): For each timescale (j, k), the divergence index

$$\Delta(j,k) = \left[S^1_j(k/T) - S^2_j(k/T) \right]^2,$$

measures the ability to separate the groups. These divergence values are then ordered and only the top pre-specified proportion of the coefficients need to be chosen.

2. Discrimination based on Kullback-Leibler (KL) discriminant crite-rion. According to the KL criterion, $\mathbf{W} = \{W_{t,T}\}_{t=1,\ldots,T}, T \geq 1\}$, is assigned to the first group if

$$\hat{I}(W:2) - \hat{I}(W:1) \geq 0 \tag{16}$$

where

$$\hat{I}(W:i) = \frac{1}{2} \sum_{j,k \in \Theta} \left[\frac{S_j^W(k/T)}{S_j^i(k/T)} + \frac{S_j^i(k/T)}{S_j^W(k/T)} - 1 \right] \quad \text{for } i = 1, 2$$

3. Discrimination based on Chernoff (CH) criterion. According to the CH criterion, $\mathbf{W} = \{W_{t,T}\}_{t=1,\ldots,T}, T \geq 1\}$, is assigned to first group if

$$\hat{Q}(W:2) - \hat{Q}(W:1) \geq 0 \tag{17}$$

where

$$\hat{Q}(W:i) = \frac{1}{2T} \left\{ \sum_{(j,k) \in \Theta} \left[\ln \left[\frac{(1-h)S_j^i(k/T) + hS_j^W(k/T)}{S_j^i(k/T)} \right] - h \ln \left[\frac{S_j^W(k/T)}{S_j^i(k/T)} \right] \right] \right\}$$

for $i = 1, 2$.

In practice, the spectra $S_j^i(z)$ in (16) and (17) are replaced by the empirical wavelet spectra, and are averaged across time series replicates. The empirical wavelet periodogram for r-th series in group i is denoted by $L_j^{i,r}(k/T)$. We then replace $S_j^i(z)$ in (16) and (17) by its estimate $\hat{S}_j^i(k/T) = \frac{1}{n_i} \sum_{r=1}^{n_i} L_j^{i,r}(k/T)$. As in the stationary case the empirical wavelet periodogram is an asymptotically unbiased but inconsistent estimator for wavelet spectra and needs to be smoothed. In practice, a smoothed version of periodogram is used. According to the KL criterion, $\mathbf{W}$ is assigned to the first group if

$$\sum_{(j,k) \in \Theta} \left\{ L_j^W(k/T) \left[\frac{1}{\hat{S}_j^2(k/T)} - \frac{1}{\hat{S}_j^1(k/T)} \right] + \ln \left[\frac{\hat{S}_j^2(k/T)}{\hat{S}_j^1(k/T)} \right] \right\} \geq 0 \tag{18}$$

and to second group otherwise. According to the CH criterion, $\mathbf{W}$ is assigned to first group if

$$\sum_{(j,k) \in \Theta} \left[\ln \frac{(1-h)\hat{S}_j^2(k/T) + hL_j^W(k/T)}{(1-h)\hat{S}_j^1(k/T) + hL_j^W(k/T)} \right] + (1-h) \ln \left[\frac{\hat{S}_j^1(k/T)}{\hat{S}_j^2(k/T)} \right] \geq 0, \tag{19}$$

and to second group otherwise.

5. Numerical Study

In this section in order to check the performance of the discriminant criteria obtained in the previous section, a numerical study is conducted. Various kinds of time series including the stationary and different kinds of non-stationary time series were investigated. The results approve the ability of the discriminant criteria (16) and (17). Nevertheless, the simulation study done by Fryzlewicz & Ombao (2009) is repeated in order to compare the results. This helps in comparing the discriminate criterions with the other advanced and useful methods in discriminating time series which were proposed by other researchers. These methods are: the SST method, which is proposed by Shumway (2003) and Sakiyama & Taniguchi (2004); the SLEX method, which is proposed by Huang et al. (2004); and the LSW method, which is proposed by Fryzlewicz & Ombao (2009). Fryzlewicz & Ombao (2009) considered four cases:

Case 1 for abruptly changing parameters and relatively large disparity between groups.

Case 2 for abruptly changing parameters and small disparity between groups.

Case 3 for smoothly changing parameters and relatively large disparity between groups.

Case 4 for abruptly changing parameters and small disparity between groups.

The n-th time series from group $g(g = 1, 2)$, denoted by $X_{n,t}^g$, is generated from the process defined by

$$X_{t,T}^g = \phi_1^g(t)X_{n,t-1}^g + \phi_2^g X_{n,t-2}^g + \epsilon_{n,t}^g$$

where

$$\epsilon_{n,t}^g \sim^{iid} N(0,1)$$

In Case 1, $\phi_1^1(t) = \begin{cases} 0.8 & \text{t=1,\ldots, 53} \\ -0.9 & \text{t=54,\ldots, 128} \\ 0.8 & \text{t=129,\ldots,256} \end{cases}$ and $\phi_1^2(t) = \begin{cases} 0.8 & \text{t=1,\ldots, 53} \\ 0.6 & \text{t=54,\ldots, 128} \\ 0.8 & \text{t=129,\ldots,256} \end{cases}$

and $\phi_2^g(t) = -0.81, t = 1, \ldots, 256$ for both groups $g = 1, 2$.

In Case 2, $\phi_1^1(t) = \begin{cases} 0.8 & \text{t=1,\ldots, 53} \\ -0.9 & \text{t=54,\ldots, 128} \\ 0.8 & \text{t=129,\ldots,256} \end{cases}$ and $\phi_1^2(t) = \begin{cases} 0.8 & \text{t=1,\ldots, 53} \\ 0.6 & \text{t=54,\ldots, 128} \\ 0.8 & \text{t=129,\ldots,256} \end{cases}$

and $\phi_2^g(t) = -0.81, t = 1, \ldots, 256$ for both groups $g = 1, 2$.

In Case 3, $\phi_1^1(t) = -0.8[1-0.7\cos(\pi t/T)]$ and $\phi_1^2(t) = -0.8[1-0.001\cos(\pi t/T)]$, and $\phi_2^g(t) = -0.81, t = 1, \ldots, 256$ for both groups $g = 1, 2$.

In Case 4, $\phi_1^1(t) = -0.8[1 - 0.7\cos(\pi t/T)]$ and $\phi_1^2(t) = -0.8[1 - 0.1\cos(\pi t/T)]$ and $\phi_2^g(t) = -0.81, t = 1, \ldots, 256$ for both groups $g = 1, 2$.
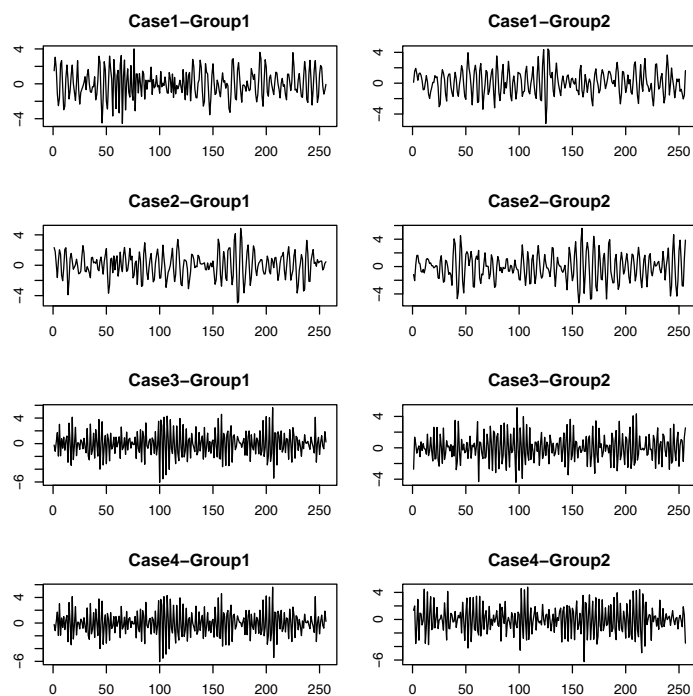
One sample of each case is shown in Figure 1.

FIGURE 1: Plot of a simulated time series for each of the cases in both groups, $g = 1, 2$.

In all cases, each of the two groups consisted of $N = 8$ time series, and each time series has length $T = 256$. Fifty test time series each of length $T = 256$ were then simulated from each group 1 and group 2. The testing time series were independent of the training data. The subset Θ of coefficients, were selected according to the top $p = 10\%$ of the timescale coefficients. Then each test time series was allocated to group 1 or 2 according to discriminant criterions (18) and (19). Daubechies (1992) identifies the Extremal Phase family: a collection of orthonormal wavelet bases possessing different degrees of smoothness and number of vanishing moments. This family of bases is indexed by the number of vanishing moments and the Haar basis is its zeroth member (Fryzlewicz 2003). Nason et al. (2000) have shown that, the J-dimensional matrix, in (12), becomes 'more' diagonal as the number of vanishing moments of the underlying wavelet increases. In all the simulations, we use the Daubechies' Extremal Phase wavelet No. 4 from the wavethresh software package for R. However, our investigations showed that, various amounts of vanishing moments have a negligible effect on the classification problem. We also eliminated the coarest scale in all calculations.

Within the Chernoff discriminate criterion, different amounts of the h parameter were examined and it was shown that small amounts lead to better results. Therefore, the amount 0.1 was chosen for the h parameter. In practice however, there would be rare cases in which the inputs of the logarithm statement became

negative in the discriminate criterions (16) and (17). These aspects were omitted from the calculations. The results are compared with three other approaches which include: the ST method, the SLEX method and the LSW method. The results are presented in Table 1.

TABLE 1: The corrected classification rate for four case models using different criterions.

Method	Case 1	Case 2	Case 3	Case 4
CH	99.5	84	95	92
KL	97	87	91	92
SST	100	69	100	96
SLEX	92	65	74	64
LSW	100	89	100	92

As can be seen in Table 1, in case 1, the SLEX method, with either finest level $J = 3$ or 4, and the LSW method, applied with the top $p = 25\%$ of timescale coefficients, perform perfectly. Other methods also work well. In case 2, the best performance was achieved by the LSW method (89%). Also, it was often the case that the CH and KL methods have accepted results with KL having slightly better result than the CH. In this case SST and the SLEX methods are not so good as other methods. It should be mentioned that case 3 is results are good, except for the SLEX method, which has a 74% correct classification rate. In this case SST and LSW have perfect performance. The CH method has slightly better results than KL method.

The results of KL, CH and LSW in case 4 are the same (92% correct classification). The best result belonged to the SST method, which has a 96% corrected classification rate. The SLEX method is the worst.

It should be mentioned that in all four methods, the conditions considered for the two groups including the number group members, the length of the time series and their models, were difficult and had complicated conditions. In general, KL and CH have shown a relatively good performance in all four cases. Meanwhile, it is been noticed that when the face the real data, the Kullback - Leibler and Chernoff criterion act much better than the LSW method.

6. Discrimination of Real Data

6.1. Discrimination of Seismic Signals

Discriminating between nuclear explosions and earthquakes is a problem of critical importance to monitor a comprehensive test-ban treaty. A dataset constructed by Blandford (1993) that comprises regional (100-2,000 km) recordings of several typical Scandinavia is used in this study. A list of these events, including eight earthquakes and eight explosions and an event of uncertain origin located in the Novaya Zemlya region of Russia, was given by Kakizawa et al. (1998). The problem was discussed in detail by (Shumway & Stoffer 2011, chap. 5) and the data is available online at `http://lib.stat.cmu.edu/general/tsa2/`. Each

Earthquake and Explosion record is actually composed of two phases: P-phase and S-phase, with 1024 points in each phase. (see Figure 2, for an earthquake, an explosion and the Novaya Zemlya (NZ) event.)
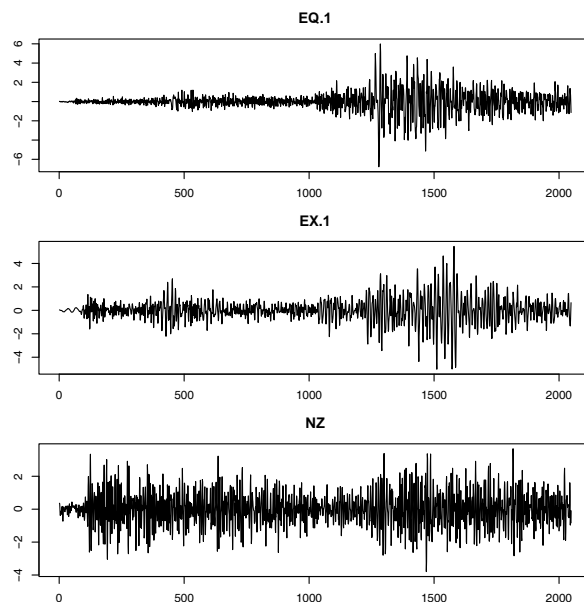

FIGURE 2: An earthquake, an explosion and the NZ event.

The results for discrimination are seen in Table 2. It should be noted that in the CH method we use $h = 0.05$ since it became obvious that a small value gave better results. To determine the rate of correct classification, we used the leave-One-Out method. The discrimination results are good, in that, in the KL method, 14 out of the 16 seismic signals were correctly classified for $p = 10\%$ and in he CH method, 13 out of the 16 seismic signals were correctly classified for $p = 10\%$ and $h = 0.05$. In our investigation, it became obvious that, for seismic data, KL discriminant criteria, when finest scale, eliminated leads to get a better result. By eliminating finest scale, 15 out of the 16 seismic signals were correctly classified in the KL method. In the LSW method, when applied with the top $p = 0.1k$, $k = 1, \ldots, 10$, of timescale coefficients, only 14 out of the 16 seismic signals were correctly classified.

Moreover, in both te KL and CH methods, the Novaya Zemlya event is classified as an explosion, which is consistent with the findings in Huang et al. (2004) and Fryzlewicz & Ombao (2009).

TABLE 2: The number of misclassifications using a leave-one-out cross validation procedure in the discrimination of seismic signals based on criterion (16) and (17).

$\Theta \longrightarrow$	scale 1-10	scale 1-9
Kullback - Leibler	2	1
Chernoff ($h = 0.05$)	3	3

6.2. Discrimination of Population Data

The population dataset was a collection of time-series representing the population estimates from 1900-1999 in 20 us states, and the data is available online from `http://www.census.gov/population/www/estimates/ststts.html`. Some of these time-series had an exponentially increasing trend while others had a stabilizing trend. The 20 states were partitioned into two groups based on their trends: group 1 consisted of CA, CO, FL, GA, MD, NC, SC, TN, TX, VA, and WA and had the exponentially increasing trend, while group 2 consisted of IL, MA, MI, NJ, NY, OK, PA, ND, and SD and had a stabilizing trend (Kalpakis, Gada & Puttagunta 2001).

Since the time series length is 99, we use a novel approach to use the information as much as possible. We partitioning the time series into 3 parts: the first part with a length of 64, consists of 1900-1963, the second part with a length of 32, consists of 1964-1995 and finally the third part with a length of 4 consists of 1996-1999, which was eliminated. We then standardized the first and second parts of the data sets. In order to under take discrimination of this data set, we used the sum of our proposed KL and CH criteria and Euclidian distance (which is proposed by Fryzlewicz & Ombao (2009)) in two parts. All three methods applied with the top $p = 10\%$ of timescale coefficients and the coarsest scale were eliminated. Once again within the CH criterion, different amounts of the h parameter were examined and it was shown that small amounts lead to better results. Therefore, the amount 0.1 was chosen for the h parameter. In the leave-One-Out method, all time series were correctly classified by the CH and KL criterions. However, in the LSW method, 18 out of 20 time series were correctly classified, the NC and MI were misclassified.

References

Blandford, R. (1993), Discrimination of earthquakes and explosions. AFTAC-TR-93-044 HQ, Air Force Technical Applications Center, Patrick Air Force Base, Florida.

Chernoff, H. (1952), 'A measure of asymptotic efficiency for tests of a hypothesis based on the sum of the observations', *The Annals of Mathematical Statistics* **23**(4), 573–578.

Dahlhaus, R. (1997), 'Fitting time series models to non stationary processes', *The Annals of Statistics* **25**, 1–37.

Dargahi-Noubary, G. R. & Laycock, P. J. (1981), 'Spectral ratio discriminants and information theory', *Journal of Time Series Analysis* **2**, 71–86.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

Fryzlewicz, P. (2003), Wavelet Techniques for Time Series and Poisson Data, Phd thesis, University of Bristol.

Fryzlewicz, P. & Ombao, H. (2009), 'Consistent classification of non stationary time series using stochastic wavelet representations', *Journl of American Statistical Association* **104**, 299–312.

Gersch, W., Yeager, C. L., Diamond, S. L., Spire, J. P., Gerry, Z. M. & Gevins, A. H. (1975), Automated analysis of the electrical activity of the human brain (eeg), Proceedings of the IEEE 63, University of California Medical Center.

Gersch, W. & Yonemoto, J. (1977), 'Automatic classification of multivariate eeg, using an amount of information measure and the eigenvalues of parametric time series model features', *Compute. Biomed. Reserch* **10**, 297–316.

Huang, H. S., Ombao, H. & Stoffer, D. (2004), 'Discrimination and classification of non-stationary time series using the slex model', *Journl of American Statistical Association* **99**, 763–774.

Kakizawa, Y., Shumway, R. & Taniguchi, M. (1998), 'Discrimination and clustering for multivariate time series', *Journl of American Statistical Association* **93**, 328–340.

Kalpakis, K., Gada, D. & Puttagunta, V. (2001), Distance measures for effective clustering of arima time-series, *in* 'Proceedings of the IEEE International Conference on Data Mining', San Jose, California, pp. 273–280.

Kullback, S. (1978), *Information Theory and Statistics*, Dover Publications, Inc, New York.

Kullback, S. & Leibler, R. A. (1951), 'On information and sufficiency', *The Annals Mathematical Statistics* **22**, 79–86.

Maharaj, E. A. & Alonso, A. M. (2007), 'Discrimination of locally stationary time series using wavelets', *Computational Statistics and Data Analysis* **52**, 879–895.

Nason, G. P., von Sachs, R. & Kroisandt, G. (2000), 'Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum', *Journal of Royal Statisticl Society, Series B* **62**, 271–292.

Ombao, H., Raz, J., von Saches, R. & Guo, W. (2002), 'The SLEX model of a nonstationary random process', *Annals of the Institute of Statistical Mathematics* **54**, 171–200.

Ombao, H., Raz, J., von Saches, R. & Malow, B. (2001), 'Automatic statistical analysis of bivariate non stationary time series', *Journal of American Statistical Association* **96**, 543–560.

Parzen, E. (1990), *Time Series, Statistics and Information*, IMA Preprint Series 663, Institute for Mathematics and Its Applications, University of Minnesota.

Sakiyama, K. & Taniguchi, M. (2004), 'Discriminant analysis for locally stationary processes', *Journal of Multivariate Analysis* **90**, 282–300.

Shumway, R. H. (2003), 'Time-frequency clustering and discriminant analysis', *Statistics and Probability Letters* **63**, 307–314.

Shumway, R. H. & Stoffer, D. S. (2011), *Time Series Analysis and Applications*, 3 edn, v, University of Minnesota.

Shumway, R. H. & Unger, A. N. (1974), 'Linear discriminant function for stationary time series', *Journl of American Statistical Association* **69**, 948–956.

Zhang, G. & Taniguchi, M. (1992), 'Discriminant analysis for stationary vector series', *ournal of Time Series Analysis* **15**, 117–126.